Probabilistic Programming Languages (PPL)

MPRI 2.40 - 2024/2025

Higher-Order Semantics

Christine Tasson

October, the 3rd

Today's schedule

Semantics, Higher Order and discrete distributions

- Syntax of a functional discrete PPL
- Discrete Markov Processes and Stochastic Matrices
- Probabilistic Coherent Spaces

Practical Session

- Modeling
- Computing semantics

Semantics, Higher-Order and continuous distribution

- Continuous Markov Processes and Kernels
- The problem with higher-order (Aumman's Lemma)
- Quasi Borel Spaces and ICones

Higher-order PPL with discrete distributions

Functional language

A Call-by-value probabilistic λ-calculus

Ground types: $b ::= \mathrm{unit} \mid \mathrm{bool} \mid \mathrm{int} \mid b \times b$

Types: $t ::= b\ \mathrm{dist} \mid t \otimes t \mid t \to t$

Constants: $c ::= () \mid \mathrm{true} \mid \mathrm{false} \mid \mathrm{n} \mid (c, c) \mid f(c, \ldots, c)$ where f is a ground operator

Values: $v ::= \mathrm{Dirac}\ c \mid \mathrm{Bernoulli\ r} \mid \mathrm{Dice\ n} \mid \lambda x.e \mid \mathrm{fix} x.e$

Terms: $e ::= v \mid x \mid (e, e) \mid (v)w \mid \mathrm{if}\ v\colon e\ \mathrm{else}\colon e \mid$ let $x =$ Sample e in e

Contexts: $G ::= 1 \mid G_0, x : t$ where $x \notin G_0$

$\boxed{G \vdash_{\mathtt{k}} e : t}$ Type System

Deterministic ($\mathtt{k} = \mathtt{d}$)

$$\frac{}{G \vdash_{\mathtt{d}} () : \mathsf{unit}} \qquad \frac{}{G \vdash_{\mathtt{d}} \underline{\mathtt{true}} : \mathsf{bool}} \qquad \frac{n \in \mathbb{N}}{G \vdash_{\mathtt{d}} \underline{n} : \mathsf{int}} \qquad \frac{f \in b_1 \times \cdots \times b_n \to b'}{G \vdash_{\mathtt{d}} f(\vec{c}) : b'} \qquad \frac{n \in \mathbb{N}}{G \vdash_{\mathtt{d}} \mathsf{Dice}\ n : \mathsf{int\ dist}}$$

$$\frac{r \in \mathbb{R}}{G \vdash_{\mathtt{d}} \mathsf{Bernoulli}\ r : \mathsf{bool\ dist}} \qquad \frac{G \vdash_{\mathtt{d}} c : b}{G \vdash_{\mathtt{d}} \mathsf{Dirac}(c) : b\ \mathsf{dist}}$$

Deterministic or Probabilistic ($k = d$ or $k = p$)

$$\frac{G \vdash_{\mathtt{k}} e_1 : t_1 \qquad G \vdash_{\mathtt{k}} e_2 : t_2}{G \vdash_{\mathtt{k}} (e_1,\ e_2) : t_1 \times t_2} \qquad \frac{G + [x : t_1] \vdash_{\mathtt{k}} e : t_2}{G \vdash_{\mathtt{k}} \lambda x.e : t_1 \to t_2} \qquad \frac{G \vdash_{\mathtt{k}} e_2 : t_1 \to t_2 \qquad G \vdash_{\mathtt{k}} e_1 : t_1}{G \vdash_{\mathtt{k}} (e_2) e_1 : t_2} \qquad \frac{G, [x : t] \vdash_{\mathtt{k}} e : t}{G \vdash_{\mathtt{k}} \mathsf{fix}\ x.e : t}$$

$$\frac{G \vdash_{\mathtt{k}} e_1 : t_1 \qquad G + [x : t] \vdash_{\mathtt{k}} e_2 : t_2}{G \vdash_{\mathtt{k}} \mathsf{let}\ x = e_1\ \mathsf{in}\ e_2 : t_2} \qquad \frac{G \vdash_{\mathtt{d}} e : \mathit{bool} \qquad G \vdash_{\mathtt{k}} e_1 : t \qquad G \vdash_{\mathtt{k}} e_2 : t}{G \vdash_{\mathtt{k}} \mathsf{if}\ e\ \mathsf{then}\ e_1\ \mathsf{else}\ e_2 : t}$$

Probabilistic ($k = p$)

$$\frac{G \vdash_{\mathtt{d}} e : t}{G \vdash_{\mathtt{p}} e : t} \qquad \frac{G \vdash_{\mathtt{d}} e : b\ \mathsf{dist}}{G \vdash_{\mathtt{p}} \mathsf{sample}(e) : b} \qquad \frac{G \vdash_{\mathtt{d}} e : \mathit{bool}}{G \vdash_{\mathtt{p}} \mathsf{assume}(P) : \mathit{unit}} \qquad \frac{G \vdash_{\mathtt{p}} e : b}{G \vdash_{\mathtt{d}} \mathsf{infer}\ e : b\ \mathsf{dist}}$$

Operational Semantics (discrete case) Labeled Transition System (LTS)

Ugo Dal Lago, On Probabilistic Lambda-Calculi, in Foundations of Probabilistic Programming, 2020

Deterministic Reduction

$$\frac{}{\text{let } x = v \text{ in } e \to_{\mathtt{d}} e[v/x]} \qquad \frac{}{(\lambda x.e)v \to_{\mathtt{d}} e[v/x]} \qquad \frac{}{\text{fix } x.e \to_{\mathtt{d}} e[\text{fix } x.e/x]}$$

Probabilistic Reduction

$$\frac{e \to_{\mathtt{d}} e'}{e \xrightarrow{1} e'} \qquad \frac{e \to_{\mathtt{d}} e'}{\text{sample } (e) \xrightarrow{1} \text{sample } (e')} \qquad \frac{\forall i \in \{1, \dots, n\}}{\text{sample } (\text{Dice } n) \xrightarrow{\frac{1}{n}} \underline{i}}$$

$$\frac{}{\text{sample } (\text{Bernoulli } r) \xrightarrow{r} \underline{\mathtt{true}}} \qquad \frac{}{\text{sample } (\text{Bernoulli } r) \xrightarrow{1-r} \underline{\mathtt{false}}} \qquad \frac{}{\text{sample } (\text{Dirac } v) \xrightarrow{1} v}$$

$$\frac{e \xrightarrow{r} e'}{\mathrm{f}(e) \xrightarrow{r} \mathrm{f}(e')} \qquad \frac{e_2 \xrightarrow{r} e_2'}{(e_2)e_1 \xrightarrow{r} (e_2')e_1} \qquad \frac{e_1 \xrightarrow{r} e_1'}{(e_2)e_1 \xrightarrow{r} (e_2)e_1'} \qquad \frac{e_1 \xrightarrow{r} e_1'}{\text{let } x = e_1 \text{ in } e_2 \xrightarrow{r} \text{let } x = e_1' \text{ in } e_2}$$

Useful Notions in Probability Theory: Stochastic Matrices

Discrete measures

$x \in (\mathbb{R}^+)^{|X|}$ is a countable distribution if $\sum_{a \in A} x_a = 1$

$x \in (\mathbb{R}^+)^{|X|}$ is a countable subdistribution if $\sum_{a \in A} x_a \leq 1$

$x \in (\mathbb{R}^+)^{|X|}$ is a countable bounded measure if $\sum_{a \in A} x_a < \infty$

sMat: Objects are $(|X|, \mathscr{P}(X))$ where $\mathscr{P}(X)$ is a set of countable bounded measures over a set $|X|$ and morphisms are stochastic matrices

Stochastic Matrix: $\theta : X \rightsquigarrow Y$ is a matrix $\mathrm{sMat}(X,Y) \subset (\mathbb{R}^+)^{|X| \times |Y|}$ such that

$$\forall x \in \mathscr{P}(X)\ \forall b \in |Y|,\ \theta(x,b) = \sum_{a \in |X|} \theta_{a,b} x_a \qquad \theta \cdot x \in \mathscr{P}(Y)$$

Composition: $\theta : \mathscr{X} \rightsquigarrow \mathscr{Y}$ and $\theta' : \mathscr{Y} \rightsquigarrow \mathscr{Z}$.

$$\forall x \in \mathscr{P}(X)\ \forall c \in |\mathscr{Z}|, \quad \theta' \circ \theta\ (x,c) = \sum_{a \in |X|} \sum_{b \in |Y|} \theta'_{b,c} \theta_{a,b} x_a = \sum_{b \in |Y|} \theta'(\theta(x,b), c)$$

Operational Semantics (Discrete case)

Stochastic Matrix:

$$\mathbf{Proba}(e,e') = \begin{cases} p & \text{if } e \xrightarrow{p} e' \\ 1 & \text{if } e \text{ does not reduce and } e = e' \\ 0 & \text{otherwise.} \end{cases}$$

Iterated Transition Matrix:

$\mathbf{Proba}^k(e,e')$ is the probability that e reduces to e' in at most k steps.

$\mathbf{Proba}^\infty(e,e')$ when e' does not reduce and is normal, is the probability that e reduces to e' in any number of steps

Higher-order PPL with discrete distributions

Goal: Adequacy

Operational Semantics computes for every closed term of ground type $\vdash_{\mathrm{p}} e : b$, the probability of reduction to each potential value $\mathbf{Proba}^{\infty}(e, \_) : |b| \to \mathbb{R}^+$.

Denotational Semantics defines $\left\{\!\!\left[\Gamma \vdash_{\mathrm{p}} e : t\right]\!\!\right\} : \{\![\Gamma]\!\} \rightsquigarrow \{\![t]\!\}$ which is an invariant of probabilistic reduction and such that adequacy holds: for every ground type b,

$$\forall v \in b,\ \left\{\!\!\left[\vdash_{\mathrm{p}} e : b\right]\!\!\right\}_v = \mathbf{Proba}^{\infty}(e, v)$$

Model of Linear Logic

SMCC - Symmetric Monoidal Closed Category

Category : $\mathscr{C}$ objects + morphisms, identity and associative composition

Symmetric Monoidal : $1, A \otimes B$ + unit, associative and commutative laws

Closed: $A \multimap B, \Lambda, \mathrm{ev}$ + currying and evaluation are in tightly related, $\dfrac{A \otimes B \to C}{A \to B \multimap C}$.

Comonad

$! : \mathscr{C} \to \mathscr{C}$ a functor (action on objects and morphisms)

counit $\epsilon_A : !A \to A$ and comultiplication $\delta_A : !A \to !!A$ + diagrams.

Monoidal Strength:

natural isomorphisms $!T \xrightarrow{\sim} 1$ and $!(A \times B) \xrightarrow{\sim} !A \otimes !B$ + coherence diagrams

Commutative Comonoid

weakening $w_A : !A \to 1$ and contraction $c_A : !A \to !A \otimes !A$ + coherence diagrams

Call-By-Value in models of LL

Call-by-name, call-by-value, call-by-need, and the Linear Lambda Calculus, Maraist & al.

Interpretation of types

Ground types: $Z^* = !Z$

Function types: $(A \to B)^* = !(A^* \multimap B^*)$

Types A^* are preceded by !, thus

$$A^* \xrightarrow{\delta} !A^*$$
$$A^* \xrightarrow{c} A^* \otimes A^* \qquad A^* \xrightarrow{w} 1$$

Contexts: $(A_1, \ldots, A_n)^* = {A_1}^* \otimes \cdots \otimes {A_n}^*$

Interpretation of terms: $G \vdash e : A$ is interpreted as a morphism $e^* : G^* \to A^*$

Semantics of CBV in LL

Interpretation of terms: $G \vdash e : A$ is interpreted as a morphism $e^* : G^* \to A^*$

Variable:

$$G, x : t \vdash x : t \qquad\qquad G^* \otimes t^* \xrightarrow{w_{G^*} \otimes \mathrm{id}} 1 \otimes t^* \xrightarrow{\sim} t^*$$

Abstraction

$$\frac{G, x : t_1 \vdash e : t_2}{G \vdash \lambda x.e : t_1 \to t_2} \qquad\qquad \frac{G^* \otimes {t_1}^* \xrightarrow{e^*} {t_2}^*}{G^* \xrightarrow{\Lambda e^*} {t_1}^* \multimap {t_2}^*}$$

$$G^* \xrightarrow{\delta} !G^* \xrightarrow{!\Lambda e^*} !({t_1}^* \multimap {t_2}^*) = t_1 \to {t_2}^*$$

Application

$$\frac{G \vdash e_2 : t_1 \to t_2 \qquad G \vdash e_1 : t_1}{G \vdash (e_2)e_1 : t_2} \qquad\qquad G^* \xrightarrow{e_2} !({t_1}^* \multimap {t_2}^*) \xrightarrow{\epsilon} {t_1}^* \multimap {t_2}^* \qquad G^* \xrightarrow{e_1} {t_1}^*$$

$$G^* \xrightarrow{c} G^* \otimes G^* \to ({t_1}^* \multimap {t_2}^*) \otimes {t_1}^* \xrightarrow{ev} {t_2}^*$$

Probabilistic Coherence Spaces

A model of LL for discrete probability

Vincent Danos, Thomas Ehrhard, On probabilistic coherence spaces, 2011

PCOH

The category of Probabilistic Coherent spaces - Pcoh

Object: $(|A|, \mathrm{P}(A))$ with $|A|$ a set and $\mathrm{P}(A) \subset (\mathbb{R}^+)^{|A|}$ such that

$$\mathrm{P}(A) = \mathrm{P}(A)^{\perp\perp} \text{ where } \mathrm{P}^\perp = \{x \in (\mathbb{R}^+)^{|A|} \mid \forall x' \in \mathrm{P}\ \langle x, x' \rangle = \sum_{a \in |A|} x_a x' a \leq 1\}$$

Bounded covering $\forall a \in |A|\ (\exists x \in \mathrm{P}(A),\ x_a \neq 0)$ and $(\exists p \in \mathbb{R}^+,\ \forall x\ x_a < p)$

Examples $[\![\tau]\!] = (|\tau|, \mathrm{P}(\tau))$

$$|\texttt{unit dist}| = \{*\} \qquad \mathrm{P}(\texttt{unit dist}) = [0,1]$$

$$|\texttt{int dist}| = \mathbb{N} \qquad \mathrm{P}(\texttt{int dist}) = \left\{(x_n) \middle| \sum x_n \leq 1\right\}$$

$$|\texttt{bool dist}| = \{\texttt{t},\ \texttt{f}\} \qquad \mathrm{P}(\texttt{bool dist}) = \left\{(x_\texttt{t}, x_\texttt{f}) \middle| x_\texttt{t} + x_\texttt{f} \leq 1\right\}$$

$$|A \times B| = |A| \uplus |B| \qquad \mathrm{P}(A \times B) = \left\{(x_i)_{i \in |A| \uplus |B|} \middle| (x_i)_{i \in |A|} \in \mathrm{P}(A), (x_i)_{i \in |B|} \in \mathrm{P}(B)\right\}$$

$$[\![\text{Bernoulli p}]\!] = (p, 1-p) \qquad [\![\text{Dice n}]\!] = \left(\tfrac{1}{n}, \ldots, \tfrac{1}{n}, 0, \ldots, 0, \ldots\right)$$

PCOH and Linear coherent maps is a model of simply typed lambda-calculus

The linear category of Probabilistic Coherence Spaces

Object: $(|A|, \mathrm{P}(A))$ with $|A|$ a set and $\mathrm{P}(A) \subset (\mathbb{R}^{+})^{|A|}$ set of functions

Morphism: $f : (|A|, \mathrm{P}(A)) \to (|B|, \mathrm{P}(B))$ a matrix $(f_{(a,b)})$ indexed by $|A| \times |B|$ such that

$$\forall x \in \mathrm{P}(A), f(x) : b \mapsto \sum_{a \in |A|} x(a) f_{a,b} \in \mathrm{P}(B)$$

Examples

$$f : [\![\text{bool dist}]\!] \to [\![\text{bool dist}]\!] \text{ such that } f = \begin{array}{c c} W\backslash S & \begin{array}{cc} \mathrm{T} & \mathrm{F} \end{array} \\ \begin{array}{c} \mathrm{T} \\ \mathrm{F} \end{array} & \begin{bmatrix} 1/5 & 4/5 \\ 3/4 & 1/4 \end{bmatrix} \end{array}$$

PCOH and Linear coherent maps is a model of Linear Logic

The linear category of Probabilistic Coherence Spaces

Object: $(|A|, \mathrm{P}(A))$ with $|A|$ a set and $\mathrm{P}(A) \subset (\mathbb{R}^+)^{|A|}$ set of functions
Morphism: $f : (|A|, \mathrm{P}(A)) \to (|B|, \mathrm{P}(B))$ such that $f \cdot \mathrm{P}(A) \subseteq \mathrm{P}(B)$

Tensor product

$$|X \otimes Y| = |X| \times |Y|$$
$$\mathrm{P}(X \otimes Y) = \{x \otimes y \mid x \in \mathrm{P}(X),\ y \in \mathrm{P}(Y)\}^{\perp\perp}$$
where $(x \otimes y) : (a, b) \mapsto x(a)y(b)$

Examples of morphisms in PCOH:

Duplication: $\Delta : [\![\text{b dist}]\!] \to [\![b \times b\text{dist}]\!]$ such that $\Delta(x) = \sum_a x_a \delta_{a,a}$
Marginalization: $\text{proj} : [\![\text{b dist}]\!] \otimes [\![\text{b' dist}]\!] \to [\![\text{b dist}]\!]$ such that $\text{proj}(x \otimes y) = x$

Exponential

$$|!X| = \mathcal{M}_{\text{fin}}(|X|)$$
$$\mathrm{P}(!X) = \{x^! \mid x \in \mathrm{P}(X)\}^{\perp\perp}$$
where $x^! : m \mapsto \prod_{a \in m} x(a)^{m(a)}$

The cartesian closed category of Probabilistic Coherence spaces

The category of Probabilistic Coherence spaces and analytic maps

Object: $(|A|, \mathrm{P}(A))$ with $|A|$ a set and $\mathrm{P}(A) \subset (\mathbb{R}^+)^{|A|}$ set of functions

Morphism: $f : (|A|, \mathrm{P}(A)) \to (|B|, \mathrm{P}(B))$ a matrix $(f_{(m,b)})$ indexed by $\mathscr{M}_{\mathrm{fin}}(|A|) \times |B|$ such that

$$\forall x \in \mathrm{P}(A), f(x) : b \mapsto \sum_{m \in \mathscr{M}_{\mathrm{fin}}(|A|)} \prod_{a \in m} x(a)^{m(a)} f_{m,b} \in \mathrm{P}(B)$$

Examples

$f : [\![unit]\!] \to [\![unit]\!]$ such that $\forall x \in [0,1],\ f \cdot x = \sum_n f_n x^n \in [0,1]$

$f : [\![bool]\!] \to [\![unit]\!]$ such that $f_{(\mathtt{true}^n, *)} = 1$ otherwise $f_{m,*} = 0$, then $f \cdot (p, 1-p) = \sum_n p^n$ and $f \cdot (1,0) = 0$.

PCOH and analytic maps is a CCC (the Kleisli Category of !) CPO enriched.

It is a model of PCF = CBN lambda calculus and fixpoints

Results on Probabilistic Coherent Spaces

Ehrhard, Pagani, Tasson, Full Abstraction for Probabilistic PCF, 2015

Compositionality

$$\llbracket (e)e_2 \rrbracket_b = \llbracket e \rrbracket (\llbracket e_2 \rrbracket)_b = \sum_m \llbracket e \rrbracket_{m,b} \prod_{a \in m} \llbracket e_2 \rrbracket_a^{m(a)}$$

Invariance of the semantics

$$\llbracket e \rrbracket = \sum_{e_2} \mathbf{Proba}(e, e_2) \llbracket e_2 \rrbracket$$

Adequacy Lemma

$$\text{if } \vdash e : \mathtt{nat}, \text{ then } \mathbf{Proba}^{\infty}(e, \underline{n}) = \llbracket e \rrbracket_n$$

Full Abstraction at ground type nat

$$\llbracket e_1 \rrbracket = \llbracket e_2 \rrbracket \text{ if and only if } \mathbf{Proba}^{\infty}(C[e_1], n) \overset{\forall C[] \forall n}{=} \mathbf{Proba}^{\infty}(C[e_2], n)$$

Why you should care or not on Probabilistic Coherence Spaces

- ✓ Interpretation of probabilistic programs of discrete type (int or real) are PCOH maps, with good analytic Properties
  - Probabilistic Stable Functions on Discrete Cones are Power Series. R. Crubille, 2018
- ✓ Concrete Sandbox for getting intuitions on probabilistic programs
- Non definability: $\mathsf{Pcoh}(\mathsf{bool}, 1) = \left\{ Q \in (\mathbb{R}^+)^{\mathcal{M}_{\mathsf{fin}}(\mathtt{t},\mathtt{f})} \mid Q_{[\mathtt{t}^n,\mathtt{f}^m]} \leq \frac{(n+m)^{n+m}}{n^n m^m} \right\}$ but the greatest we can get is $\{\![e]\!\} \leq \frac{(n+m)!}{n!m!}$

  fix fun x (*$\to$*) if x then if x then f(x) else () else if x then () else f(x)

Why you should care or not on Probabilistic Coherence Spaces

- It is not computable and thus cannot be used to implement inference
  if $\vdash M : \tau$ and $\{\!\![M]\!\!\} \in \mathrm{P}(\tau)$ then $\vdash \text{infer } M : \tau\, dist$ and $(\!(\text{infer } M)\!)$ is a subprobability distribution over τ.

$$(\!(\text{infer } M)\!) = \frac{\{\!\![M]\!\!\}}{\sum_{a \in |\tau|} \{\!\![M]\!\!\}_a}$$

Scaling exact inference for discrete probabilistic programs. Holtzen & al. 2020

Practical Session

Modeling & Computing Semantics

Higher-order continuous PPL

From LTS to Stochastic Kernels

Useful Notions in Measure Theory:

Meas:

Objects are Measurable spaces: $\mathscr{X} = (|\mathscr{X}|, \Sigma_{\mathscr{X}})$

$|\mathscr{X}|$ set of outcomes of a probabilistic experience,

$\Sigma_{\mathscr{X}} \subseteq \mathscr{P}(|\mathscr{X}|)$ sigma-algebra (closed under countable unions, intersections and complement)

Example: choose uniformly two reals in $[0,1]$: $|X| = [0,1] \times [0,1]$ with the Sigma-algebra generated by $[a,b] \times [c,d]$

Morphisms are Measurable Function: $f : \mathscr{X} \to \mathscr{Y}$ is a function such that $\forall V \in \Sigma_{\mathscr{Y}},\ f^{-1}(V) \in \Sigma_{\mathscr{X}}$

Bounded **Measure:** $\mu : \Sigma_{\mathscr{X}} \to \mathbb{R}^+$ sigma-additive, that is

$\mu(\emptyset) = 0$ and $\mu(|X|) < \infty$, $\mu(\sqcup A_i) = \sum_i \mu(A_i)$, $\mu(|X| \setminus A) = \mu(|X|) - \mu(A)$

Tensored Borel Measure: $\beta \otimes \beta([a,b] \times [c,d]) = \beta([a,b]) * \beta([c,d]) = (b-a) * (d-c)$

Pushforward measure $f_* \mu \in$ **Meas**$(\mathscr{Y})$**:** $\mu \in \mathrm{Meas}(\mathscr{X})$, $f \in \mathrm{Meas}(\mathscr{X}, \mathscr{Y})$,

Useful Notions in Measure Theory: Stochastic Kernels

sKern: Objects are measurable spaces: $\mathscr{X}$ and morphisms $\mathrm{sKern}(\mathscr{X},\mathscr{Y}) = \mathrm{Meas}(\mathscr{X},\mathrm{Meas}(\mathscr{Y}))$

Stochastic Kernels: $\kappa : \mathscr{X} \rightsquigarrow \mathscr{Y}$ is a function: $\kappa : \mathscr{X} \times \Sigma_{\mathscr{Y}} \to \mathbb{R}^+$ such that

$\forall x \in \mathscr{X},\ \kappa(x,\_) : \Sigma_{\mathscr{Y}} \to \mathbb{R}^+$ is a measure

$\forall V \in \Sigma_{\mathscr{Y}},\ \kappa(\_,V) : \mathscr{X} \to \mathbb{R}^+$ is a measurable function

Composition: $\kappa : \mathscr{X} \rightsquigarrow \mathscr{Y}$ and $\kappa' : \mathscr{Y} \rightsquigarrow \mathscr{Z}$.

$$\forall x \in |\mathscr{X}|\ \forall W \in \Sigma_{\mathscr{Z}}, \quad \kappa' \circ \kappa(x)(W) = \int_{\mathscr{Y}} \kappa'(y,W)\kappa(x,dy)$$

Giry Monad: $\mathscr{G}(\mathscr{X}) = \mathrm{Meas}(\mathscr{X}), \Sigma(\{\mu \in \mathscr{G}(X) \mid \mu(U) < r\}_{\mu \in \mathscr{G}(\mathscr{X}), r \in \mathbb{R}})$:

Unit: Dirac $x \in \mathrm{Meas}(\mathscr{X}) : (\mathrm{Dirac}\ x)(U) = \delta_{x \in U}$

Bind: if $\mu \in \mathrm{Meas}(\mathscr{X})$, $\kappa \in \mathrm{Meas}(\mathscr{X},\mathrm{Meas}(\mathscr{Y}))$, then $\mu \blacksquare \kappa(U) = \int_{\mathscr{X}} \kappa(x)(U)\mu(dx)$

Operational Semantics (Continuous case)

Stochastic Kernel: **Proba** $: \Lambda^{\Gamma\vdash t} \rightsquigarrow \Lambda^{\Gamma\vdash t}$ that is $\Lambda^{\Gamma\vdash t} \times \Sigma_{\Lambda^{\Gamma\vdash t}} \to \mathbb{R}^+$

$\Lambda^{\Gamma\vdash t} = \{e \mid \Gamma \vdash e : t\}$

$\Sigma_{\Lambda^{\Gamma\vdash t}} = \{U$ measurable, i.e.$\forall n, \forall S$, $\{\vec{r}\ s.t.\ S\underline{\vec{r}} \in U\}$ meas. in $\mathbb{R}^n\}$

for all $e \in \Lambda^{\Gamma\vdash t}$, $\mathbf{Proba}(e, \_)$ is a measure;

for all $U \in \Sigma_{\Lambda^{\Gamma\vdash t}}$, $\mathbf{Proba}(\_, U)$ is a measurable function.

Iterated Stochastic Kernel:

$\mathbf{Proba}(e, U))$ is the probability to observe U after one reduction step from e.

$\mathbf{Proba}^\infty(e, U)$ is the probability to observe a normal form in U after any steps.

If $\vdash e : \mathrm{real}$, then $\mathbf{Proba}^\infty(e, \_)$ is the continuous distribution over $\mathbb{R}$ computed by e.

Operational Semantics (Continuous case)

Values: $r \mid \text{fun } x \to e$

$$\textbf{Proba}(e, U) = \begin{cases} \delta_{e'}(U) & \text{if } e \xrightarrow{1} e' \\ \delta_{e}(U) & \text{if } e \text{ does not reduce} \\ 0 & \text{otherwise.} \end{cases}$$

$$\textbf{Proba}(\texttt{sample}(\texttt{Dirac e}), U) = \delta_{e}(U)$$

$$\textbf{Proba}(\texttt{sample}(\texttt{Bernoulli r}), U) = r\,\delta_{\underline{\texttt{true}}}(U) + (1-r)\delta_{\underline{\texttt{false}}}(U)$$

$$\textbf{Proba}(\texttt{sample}(\texttt{Uniform}\ \texttt{r}_0\ \texttt{r}_1.), U) = \int_{\underline{r}\in U} \mathbb{1}_{[r_0, r_1]}(r)dr$$

Higher-order continuous PPL

Aumann's Lemma

What category for Higher-Order

Wanted A Symmetric Monodial Closed Category with:

- Objects: Meas($\mathbb{R}$) is an object and object comes with some measures and integration
- Morphims $f : t_1 \to t_2$ with some measurability such that $f \circ \delta :$ can be integrated

Aumann's Lemma $\mathbb{R} \to \mathbb{R}$ cannot be turned into a measurable space such that $\mathrm{ev} : \mathbb{R} \otimes (\mathbb{R} \to \mathbb{R}) \to \mathbb{R}$ is measurable, thus **skern** is disqualified.

Solution

Measurability problem

$$\{\!\![\, \mathtt{let}\, x = \text{sample } e\, \mathtt{in}\, e_2 \,]\!\!\} = \int_{\mathbb{R}} (f \circ \delta)(r)\ \mu(dr)$$

Use a different category including Meas(R) and extending measurability and integration to all types by a logical relation reducing to $\mathbb{R}$.

What category for Higher-Order

Wanted A Symmetric Monodial Closed Category with:

Objects: Meas($\mathbb{R}$) is an object and object comes with some measures and integration

Morphims $f : t_1 \to t_2$ with some measurability such that $f \circ \delta :$ can be integrated

Aumann's Lemma $\mathbb{R} \to \mathbb{R}$ cannot be turned into a measurable space such that $\mathrm{ev} : \mathbb{R} \otimes (\mathbb{R} \to \mathbb{R}) \to \mathbb{R}$ is measurable, thus **skern** is disqualified.

Solution

Measurability problem

$$\llbracket \mathtt{let}\, x = \text{ sample } e \,\mathtt{in}\, e_2 \rrbracket = \int_{t_1} (f \circ \delta)(r) \; \mu(dr)$$

Use a different category including Meas(R) and extending measurability and integration to all types by a logical relation reducing to $\mathbb{R}$.

Higher-order continuous PPL

Aumann's Lemma : sKern is symmetric monoidal but not closed

sKern is symmetric monoidal but not closed.

By Contradiction: Assume that evaluation $\forall X, Z, \mathrm{ev}: Z^X \otimes X \to Z$ is measurable for every X, Z.

Measurable spaces X is $\mathbb{R}$ endowed with the σ-algebra $\Sigma_X = \mathscr{P}(X)$ of all subparts and Y is $\mathbb{R}$ endowed with the σ-algebra countable-cocountable generated by countable parts and parts whose complement is countable (closed under countable unions and countable intersections).

Diagonal function: $h: \begin{cases} (\mathbb{R} \times \mathbb{R}, \mathscr{P}(\mathbb{R}) \otimes \mathscr{C}(\mathbb{R})) & \to \{0,1\} \\ (x, y) & \mapsto 1 \text{ if } x = y \\ & \mapsto 0 \text{ otherwise} \end{cases},$

$\Lambda(h): (\mathbb{R}, \mathscr{P}(\mathbb{R})) \to (\{0,1\}^{\mathbb{R}}, \Sigma_{2^Y})$ is **measurable**

$h = \mathrm{ev} \circ \Lambda(h)$ is **measurable** since it is the composite of measurable functions.

$\Delta = \{(x, y) \in \mathbb{R}^2 \mid x = y\} = h^{-1}(1)$ is measurable in $\mathscr{P}(\mathbb{R}) \otimes \mathscr{C}(\mathbb{R})$.

sKern is symmetric monoidal but not closed.

By Contradiction: Assume that evaluation $\forall X, Z, \mathrm{ev} : Z^X \otimes X \to Z$ is measurable for every X, Z.

Then $\Delta = \{(x,y) \in \mathbb{R}^2 \mid x = y\} = h^{-1}(\{1\})$ is **measurable** in $\mathscr{P}(\mathbb{R}) \otimes \mathscr{C}(\mathbb{R})$.

Proposition: Si $W \in \mathscr{P}(\mathbb{R}) \otimes \mathscr{C}(\mathbb{R})$, then, there is $B \subseteq \mathbb{R}$ dénombrable such that

If there is $(x,y) \in W$ such that $y \notin B$, then $\forall z \notin B$, $(x,z) \in W$.

Proof: it is satisfied by all base measurable sets and closed by countable union and countable intersection.

Remark: Δ satisfies this property, let B be countable Since B is countable, there is $(x,x) \in \Delta$ such that $x \notin B$.
Since B is countable, there is $z \notin B$ and $z \neq x$, thus $(x,z) \in \Delta$ and $(x,z) \notin \Delta$.

This is a **contraditon**

The Measurability Problem

Semantics framework

Type $\texttt{real}$ is interpreted as $[\![\texttt{real}]\!] = \mathrm{Meas}(\mathbb{R})$,

Closed term $[\![\vdash e : \texttt{real}]\!]$ as a measure μ and

Term $[\![x : \texttt{real} \vdash e_2 : \texttt{real}]\!]$ as a morphism $f : \mathrm{Meas}(\mathbb{R}) \to \{\!|\texttt{real}|\!\}$.

$$\{\!|\texttt{let}\, x = \text{ sample } e\, \texttt{in}\, e_2 |\!\} = \int_{\mathbb{R}} (f \circ \delta)(r)\ \mu(dr)$$

The Measurability Problem

Semantics framework

Type $\texttt{real}$ is interpreted as $[\![\texttt{real}]\!] = \mathrm{Meas}(\mathbb{R})$,

Closed term $[\![\vdash e : \texttt{real}]\!]$ as a measure μ and

Term $[\![x : \texttt{real} \vdash e_2 : \texttt{real}]\!]$ as a morphism $f : \mathrm{Meas}(\mathbb{R}) \to \{\![\texttt{real}]\!\}$.

What if $\vdash e : t_1$ and $x : t_1 \vdash e_2 : t_2$?

$$\{\![\texttt{let}\, x = \text{sample } e\, \texttt{in}\, e_2]\!\} = \int_{t_1} (f \circ \delta)(r)\ \mu(dr)$$

Two CPO-enriched CCC implementing this idea

QBS based on presheaves over Meas

- *A Convenient Category for Higher-Order Probability Theory, Chris Heunen, Ohad Kammar, Sam Staton, Hongseok Yang, LICS2017*
- *A Domain Theory for Statistical Probabilistic Programming Matthijs Vákár, Ohad Kammar, Sam Staton, POPL2019*

ICONES based on cones

- *Thomas Ehrhard, Michele Pagani, Christine Tasson. Measurable Cones and Stable, Measurable Functions POPL 2018.*
- *Thomas Ehrhard, Guillaume Geoffroy. Integration in cones. 2023. Techincal report*

Semantics of Probabilistic Programming

Quasi Borel Spaces

📕 *Ohad Kammar Tutorial*

QBS - Definition

Quasi Borel Space $X = (|X|, \mathscr{R}(X))$ such that

Samples: $|X|$ is the sample set

Random elements: $\mathscr{R}(X) \subseteq \mathbb{R} \to |X|$ Paths

Constants: if $x \in |X|$, then $\lambda r.x \in \mathscr{R}(X)$

Precomposition: if $\alpha \in \mathscr{R}(X)$ and $\varphi : \mathbb{R} \to \mathbb{R}$ measurable, then $\varphi \circ \alpha \in \mathscr{R}(X)$.

Recombination: if $\alpha \in \mathscr{R}(X)^{\mathbb{N}}$ and $\mathbb{R} = \uplus A_n$ and A_n measurable,
then $\lambda r.\alpha_n(r)$ (if $r \in A_n$) $\in \mathbb{R}(X)$

Examples

Measurable spaces: if (X, Σ) is a measurable space, then $(X, \mathrm{Meas}(\mathbb{R}, X))$ is a QBS

$[\![\text{real}]\!]$: $(\mathbb{R}, \mathrm{Meas}(\mathbb{R}, \mathbb{R}))$

$[\![\text{int}]\!]$: $(\mathbb{N}, \mathrm{Meas}(\mathbb{R}, \mathbb{N}))$

QBS - Definition

Quasi Borel Space $X = (|X|, \mathscr{R}(X))$ such that

Samples: $|X|$ is the sample set

Random elements: $\mathscr{R}(X) \subseteq \mathbb{R} \to |X|$ Paths

Constants: if $x \in |X|$, then $\lambda r.x \in \mathscr{R}(X)$

Precomposition: if $\alpha \in \mathscr{R}(X)$ and $\varphi : \mathbb{R} \to \mathbb{R}$ measurable, then $\varphi \circ \alpha \in \mathscr{R}(X)$.

Recombination: if $\alpha \in \mathscr{R}(X)^{\mathbb{N}}$ and $\mathbb{R} = \uplus A_n$ and A_n measurable,

then $\lambda r.\alpha_n(r)$ (if $r \in A_n$) $\in \mathbb{R}(X)$

Morphisms: $\mathrm{QBS}(X, Y)$

Function $f : |X| \to |Y|$ preserving random elements:

$$\text{If } \alpha \in \mathscr{R}(X), \text{ then } f \circ \alpha \in \mathscr{R}(Y)$$

notice that $\mathscr{R}(X) = \mathrm{QBS}(\mathbb{R}, X)$

QBS - Properties

QBS is a CCC

Cartesian:

$|\top| = \{*\}$ and $\mathscr{R}(\top) = \mathrm{Meas}(\mathbb{R}, \{*\})$

$|X \times Y| = |X| \times |Y|$ and

$\mathscr{R}(X \times Y) = \{\lambda r.(\alpha(r), \beta(r)) \mid \alpha \in \mathscr{R}(X), \beta \in \mathscr{R}(Y)\}$

Closed:

$\left|Y^X\right| = \mathbf{QBS}(X, Y)$ and

$\mathscr{R}(Y^X) = \{\alpha : \mathbb{R} \to Y^X \mid \lambda(r,x).\alpha(r)(x) \in \mathbf{QBS}(\mathbb{R} \times X \to Y)\}$

$ev : Y^X \times X \to Y$ is $ev(f,x) = f(x)$

Limits: Coproducts, Quotients, ... as in Sets

QBS is a **conservative extension** of Standard Borel Sets

For any (X, Σ) in Meas, $(X, \mathrm{Meas}(\mathbb{R}, X))$ is in QBS.

If (X_1, Σ_1) and (X_2, Σ_2) are in Meas, then

$$\mathrm{Meas}((X_1, \Sigma_1), (X_2, \Sigma_2)) = \mathrm{QBS}((X_1, \mathrm{Meas}(\mathbb{R}, X_1)), (X_2, \mathrm{Meas}(\mathbb{R}, X_2)))$$

Interpreting Let in QBS

Measure on a QBS $X = (|X|, \mathscr{R}(X))$

a measure over X is a pair of a measure μ over $\mathbb{R}^p$ and a path $\alpha \in \mathrm{QBS}(\mathbb{R}^p, X)$

for any QBS morphism $f : X \to Y$, the pair μ and $f \circ \alpha$ is a measure

up to isomorphisms

Integration let $f \in \mathrm{QBS}(X, \mathbb{R})$ and $[\mu, \alpha]$ a measure on X

$$\int_X f(x)[\mu, \alpha](dx) = \int_{\mathbb{R}} f \circ \alpha(r)\mu(dr)$$

Integration in QBS X boils down to intergration in $\mathbb{R}$.

Example - Linear Regression

```
def model() =
  m = sample(Gaussian(0, 2))
  b = sample(Gaussian(0, 2))
  f = lambda x: m * x + b
  return f
s = sample(infer(model))(4)
```

Measure space: $\mathbb{R}^2$ with borelians **Probability** $\mathbb{P}:\ m, b \sim \mathcal{N}(0,2) \otimes \mathcal{N}(0,2)$

Random variable: $\alpha : (m, b) \mapsto \lambda x. m * x + b$

Distribution: $(\!| model |\!) = [\alpha, \mu]$

$$\begin{aligned} [\![\text{sample}(\text{infer}(\text{model}))(4)]\!] &= \int_{\mathbb{R}^{\mathbb{R}}} f(4)[\alpha, \mu](df) \\ &= \int_{\mathbb{R}^2} (4 * m + b) \mathcal{N}(0,2)(dm) \mathcal{N}(0,2)(db) \end{aligned}$$

Conclusion

Skills and knowledge

Take home

Modeling and Semantics of PPL

Semantics of IMP and pPCF using Markov Processes

Inference is **intractable** (even in discrete Bayesian Networks)

Heuritics for efficient exact inference (in practical case)

Personal Homework

Quasi-Borel-Spaces and PPL

30% final mark

Date: 14/11/24 by email to christine.tasson@lip6.fr

Next

Build Your Own PPL

Give its semantics

Enrich its inference algorithm

Use static analysis to validate an SVI algorithm